DOCUMENT RESUME

ED 216 043 TM 820 206

AUTHOR Ligon, Glynn; Matter, Kevin
TITLE Anomalies in Achievement Analyses.
INSTITUTION Austin Independent School District, Tex.
REPORT NO AISD-81-60
PUB DATE 81
NOTE 15p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Rating; *Achievement Tests; Elementary Secondary Education; *Scores; Student Evaluation; Student Improvement; *Test Interpretation; Test Use
IDENTIFIERS Austin Independent School District TX

ABSTRACT

Six anomalies in achievement test scores encountered by the Austin Independent School District are described. These include crossing gaps with uninterpolated medians; total group median declines while all subgroups' medians rise; outlying total percentiles; percentile and grade equivalent growth antithesis; same grade equivalent earning a different percentile in each content area; and the median does not represent any group. Evaluators and researchers must know how to distinguish real achievement gains from artifactual gains resulting from anomalies such as those discussed in this paper. It is necessary to determine when an inconsistency is an error and when it is an explainable anomaly. When interpreting achievement test scores, interaction of types of scores such as percentiles and grade equivalents, shifts in student demographics, and non-normal distributions within groups being tested need to be carefully considered. The factors causing the anomalies and possible solutions are discussed. (DWH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED216043

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
F. Holley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# Anomalies in Achievement Analyses

GLYNN LIGON, Ph.D.
KEVIN MATTER
Austin Independent School District

Maybe we missed a few classes in our graduate statistics courses, or possibly our problems are not technical enough to merit journal articles. At any rate, we were temporarily stunned when unexpected anomalies and mystifying inconsistencies began to haunt our reporting of achievement test scores. We also had been collecting a list of questions which often confused teachers and other school staff. This paper pulls these anomalies and questions together to serve as a reference for anyone who reports achievement test results.

- Can each ethnic group gain more in a year than does the total group combined?
- Can each ethnic group gain while the total group's median declines?
- Can a group's percentile median on each subtest be higher than the group's median on the total score?
- Can a student gain a year in grade equivalents and lose percentile points?
- Why is the same grade equivalent equal to two different percentile ranks in reading and math?
- Can the gap between two ethnic groups' achievement decrease at each individual grade level from one year to the next but continue to widen from one grade to the next?
- Can a school's median percentile misrepresent the actual student population?

This paper approaches these issues from a practitioner's perspective. As evaluators and researchers who report achievement test scores, we need to understand when an inconsistency is an error and when it is merely an explainable anomaly.

TM 820 206



Publication No. 81.60

ANOMALY 1: CROSSING GAPS WITH UNINTERPOLATED MEDIANS
(Changing Subtly by Leaps and Bounds)

When two or more subgroups or components change in a positive or negative direction from one specific point in time to another point in time, we expect the total group to reflect this change and be the sum total of the changes, or at least as large as the smallest subgroup change. For example (see Figure 1) if these three groups change by five points, we might expect the total group to change by five points.

In the case of AISD districtwide achievement results, this pattern does not follow, as we found out in the 1980-81 school year. That was the second year in which we used the Iowa Tests of Basic Skills (ITBS), so we were anxious to see how our achievement changed from the initial year of the ITBS use, 1979-80. The achievement results were particularly important because 1980-81 was the first year of large-scale, court-ordered busing for desegregation.

In AISD the junior high students are tested in February, so they were the first "test case" for us of the ITBS norms over time.

Our initial analysis of the grade 7 Reading Total (RT) results by ethnicity looked excellent in terms of percentile gains (see Figure 2). Our District RT median percentile score for Blacks rose by 7 %ile points, as did the District RT median percentile score for Hispanics. We looked at the RT score for our Anglo/Other students and found that it had risen by 4 %ile points. By this time we were ecstatic. We looked eagerly for the District score for all students tested - and found only a 2 %ile point increase. Disappointment set in. How could such a small overall increase result from such large subgroup increases?

As believers in fully checking out our numbers, we ran a frequency distribution of the scores to verify that the middle score was in fact at the median score that we had calculated. It was. The problem now became explaining these results to our School Board and the Austin public. What happened to cause this anomaly in the scores?

First, for a given test, all 99 percentile ranks may not be achievable. The gaps between achievable percentiles vary in size at different points in the distribution. Typically in the middle percentile ranges, not all percentiles are possible; while at the extremes, each percentile rank is possible. In the case of our grade 7 RT scores, a small change in raw score moved each ethnic group's uninterpolated median across a gap which was larger than the gap spanned by the change in the total group median. The net result was that our RT gains by ethnicity were impressive, using median percentiles, while our gain as a District was not as impressive.

As noted in Figure 3, the gain in terms of grade equivalent points for RT was smaller for the Anglo students than for the Total group, but the percentile gain was four points compared to two points. This change of four percentile points was the smallest positive change possible at that point on the RT percentile scale, while at the middle of the scale a positive change was limited to two percentile points.

Secondly, because the medians for each year and each group were all independently calculated, this possibility of large increases in subgroup scores and smaller increases in the total group can always exist. These independent median calculations are not direct functions of each other. The subgroup medians do not have direct influence on the total group median score. Therefore the expected relationships, as seen in the first figure, do not hold and should not be expected.

Our response to the problems encountered in this anomaly of large increases in subgroup scores but small increases in the total group scores is twofold. First, we are investigating the use of calculating an interpolated median percentile score. As we found out over the past few years, a shift in the scores of a few students by a single point can create a large difference in the median percentile point when based upon the actual middle-scoring student. If this shift is near a large gap in the percentile tables, the resulting median score may not provide the most accurate picture of districtwide achievement. An interpolated median percentile will allow for a score which, although not truly attainable, will more accurately reflect the "middle" of the score distribution. We feel this will eliminate random increases/decreases in districtwide averages, which may not be actual changes in achievement but rather artifacts of the method used to calculate the median percentile. The use of interpolated median percentile points should more accurately assess "true" changes in achievement over time.

We also plan to give more emphasis to ITBS grade equivalent scores, which were developed as an equal-interval scale. Through the use of grade equivalent scores we hope to have a better representation of the size of changes in achievement for groups in various ranges of the distribution.

| GROUP | TIME 1 | TIME 2 | CHANGE |
|---|---|---|---|
| A | 20 | 25 | +5 |
| B | 30 | 35 | +5 |
| C | 70 | 75 | +5 |
| TOTAL | 40 | 45 | +5 |

Figure 1. Common-sense relationship between subgroup and total group percentile scores over time.

| ETHNICITY | 1980 MEDIAN | 1981 MEDIAN | CHANGE |
|---|---|---|---|
| BLACK | 20 | 27 | +7 |
| HISPANIC | 23 | 30 | +7 |
| ANGLO/OTHER | 67 | 71 | +4 |
| TOTAL | 50 | 52 | +2 |

Figure 2. A comparison of median percentile scores by ethnicity. . . . ITBS, Grade 7, Reading Total, AISD.

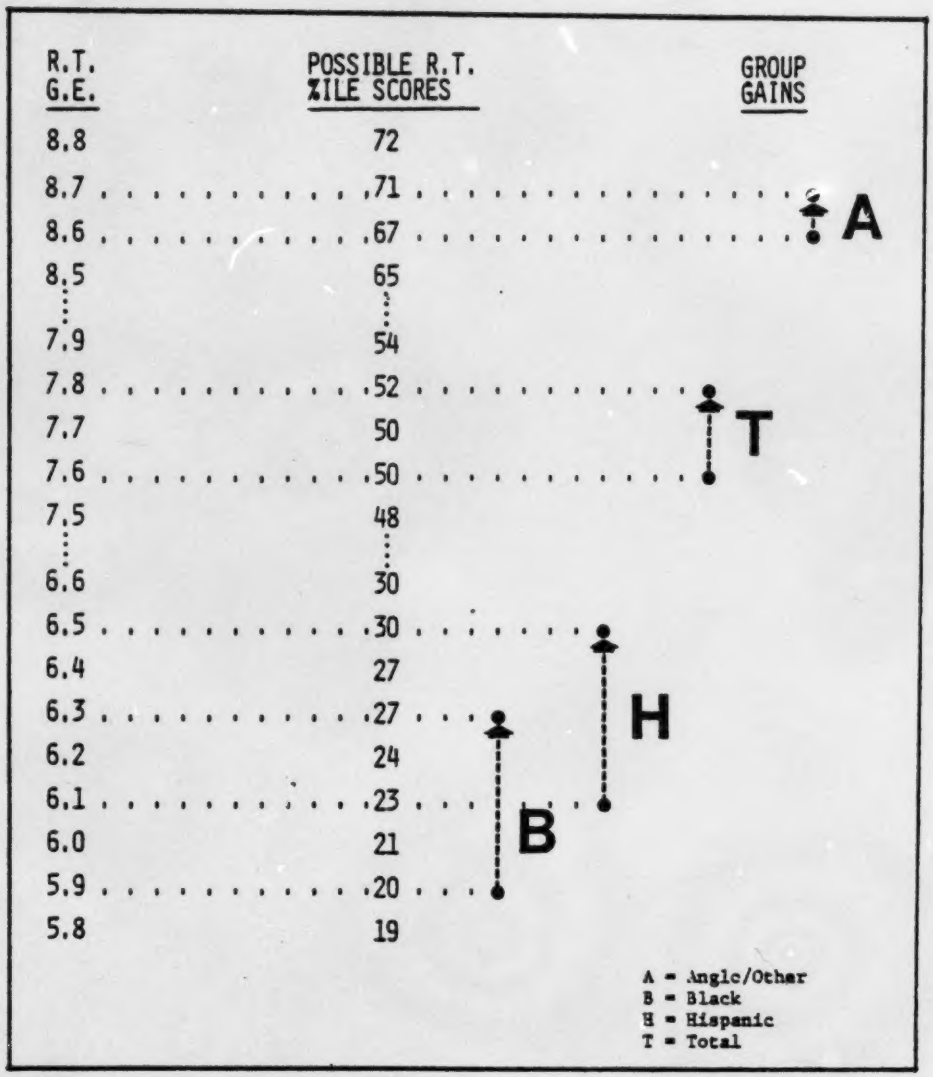


Figure 3. A comparison of median percentile and grade equivalent scores by ethnicity, ITBS, Grade 7, Reading Total, AISD.

ANOMALY 2: TOTAL GROUP'S MEDIAN DECLINES WHILE ALL SUBGROUPS' MEDIANS RISE
(How Three Positives Make a Negative)

Anomaly number two was discovered in the results of our April 1981 ITBS elementary school testing. We encountered a case where the total median percentile and grade equivalent scores actually dropped, even though the ethnic subgroup median percentile and mean grade equivalent scores rose. Again, on the surface, it seems like that is not possible. We rechecked the data, carefully multiplying the mean grade equivalent score for each ethnic group by the total number of students in that group. The results were verified - three positives did in fact make a negative. How?

As seen in Figure 4, there was a shift in the school system's population by ethnicity from 1980 to 1981. There was now a lower overall proportion of Anglo students in the District. This higher achieving group exerted less upward influence on the 1981 District total score. Even though every ethnic group's mean grade equivalent score rose, the total was influenced less by the highest achieving group.

A second factor entering into the picture was a change in the percentage of students taking the test in 1980 and in 1981, by ethnicity. An increase in the percentage of Black and Hispanic students tested in 1981 over 1980 raised the proportion of lower achieving minority students represented in the district-wide mean grade equivalent score.

With this second anomaly, the explanations of the test results are logical, and even obvious when one concentrates on the phenomena involved. But if one looks only at the numbers, the results alone, the achievement picture is puzzling.

Our response to this anomaly, a decrease in total group score while the subgroups increased, focused on estimating the impact of shifts in ethnicity and the number of students tested. We calculated an estimate of the 1981 grade equivalent scores, based upon the 1980 scores. Achievement was held constant, but we took into account the change in the number of students tested by ethnicity. These estimated 1981 grade equivalent scores were compared to the actual 1981 scores to determine the expected change in achievement which could be attributed to this shift in ethnic composition and number of students tested.

Through the use of these projected scores, AISD scores in reading would be expected to be lower in 1981 in grades 1-7 and higher in grade 8. A comparison of these projected scores with actual 1981 achievement indicated that:

- achievement improved rather than declined in grades 1, 2, and 5-7.
- achievement in grades 3 and 4 declined some, but no more than expected.
- achievement in grade 8 improved more than expected.

We are now also reporting longitudinal data for students who have been tested every year, thus, making our year-to-year comparisons on the same students rather than merely on groups whose make-up might shift.

| ETHNIC GROUP | 1980 MEAN G.E. | N | 1981 MEAN G.E. | N | CHANGE MEAN G.E. | N |
|---|---|---|---|---|---|---|
| BLACK | 3.19 | 760 | 3.30 | 757 | +.11 | -3 |
| HISPANIC | 3.33 | 1078 | 3.37 | 1108 | +.04 | +30 |
| ANGLO/OTHER | 4.46 | 2443 | 4.50 | 1917 | +.04 | -526 |
| TOTAL | 3.95 | 4281 | 3.93 | 3782 | -.02 | -499 |

Figure 4. Comparison of changes in mean grade equivalent scores from 1980 to 1981 . . . . ITBS, Reading Total, Grade 3, AISD.

## ANOMALY 3: THE OUTLYING TOTAL PERCENTILES
## (The Junior High Principal Panic)

We encountered this issue while interpreting median and quartile scores for a group of junior high principals. Although our computer programs had been checked a dozen times, the principals noticed that their totals were often noticably lower or higher than their subtests. Controlled panic began. Of course, everything was calculated correctly, and another anomaly was added to our list.

The psychometrically naive educator expects total scores to be somehow arithmetically a function of subtest scores. Unfortunately, the farther away from the 50th percentile that scores fall, the more likely that the total percentile will be farther away from 50 than are all the subtest percentiles. Figure 5 presents examples to illustrate this anomaly.

When all subtest percentiles are consistently low (or high), the percentile for the total test will usually be even lower (or higher) rather than being about midway among the subtest percentiles. The explanation for this lies in the nature of the score distributions. An individual student may score very low on one subtest but somewhat higher on the others. A pattern of very low scores on all subtests is less common and results in a total score which falls even lower in the distribution (i.e., receives a lower percentile rank).

The outlying total percentile occurs frequently with individuals' scores. However, group averages are even more prone to this phenomenon. For a group, the average subtest scores tend to be more similar than are the subtest scores for individuals. When first- and third-quartile points are reported, the outlying total percentile is quite common.

| GRADE EQUIVALENT | PERCENTILE RANK | | | | |
|---|---|---|---|---|---|
| | L1 | L2 | L3 | L4 | LT |
| **2.9** | 12 | 11 | 11 | 15 | **9** |
| **4.9** | 51 | 51 | 51 | 51 | **51** |
| **7.3** | 90 | 90 | 90 | 87 | **93** |

Figure 5. Comparison of percentile ranks associated with the same grade equivalent for low, average, and high achievers--ITBS, language tests, grade 4, spring norms.

ANOMALY 4: THE PERCENTILE AND GRADE EQUIVALENT GROWTH ANTITHESIS
(a. What Goes Up Can Also Go Down, at the Same Time.)
(b. The Hurrieder I Go, the More Behinder I Get.)

Conclusions about a student's growth in achievement may be antithetical depending upon the choice of grade equivalents or percentiles as the statistic to use in expressing gains. A student must maintain or improve a percentile rank for achievement to be considered as progressing well. In grade equivalents, a student must gain 1.0 in a calendar year to have demonstrated a year's normal growth. Unfortunately, neither represents a complete picture of achievement growth, and either alone may be misinterpreted.

a. Consider a student who scores at the 27th percentile in grade 3 and at the 28th percentile in grade 4 (Language Total, Iowa Tests of Basic Skills, 1978, spring norms).

- Did this student make better than average progress?
- Did this student make more than a year's growth?
- Is this student closer to being "on grade level" in grade 4?

The simple answer to each of these three questions is "No." Even though the student's percentile rank improved, the growth in grade equivalents was only from 3.0 to 3.9.

Consider another student who scored at the 5.1 grade equivalent level in grade 3 and at the 6.2 level in grade 4 (same test).

- Did this student's percentile rank also improve?
- Did this student make the gain that is expected of students this far above grade level?

The answer to these two questions is "No." Even though more than one year's growth was achieved, this student's percentile rank was 78 in grade 3 and 77 in grade 4.

What is also interesting is that the achievement gap between this high achiever and this low achiever increased by 0.2 grade equivalent while their percentile rank gap closed by two points.

To generalize, a student may gain more than 1.0 grade equivalent and still realize a decline in percentile rank. On the other hand, a student may gain less than 1.0 grade equivalent and realize a rise in percentile rank. Obviously, the two scales are not linked in a direct manner. Students who score below the first quartile do not have to gain 1.0 grade equivalent in a year to maintain their ranking relative to other low achievers; however, students who score above the third quartile must gain more than 1.0 grade equivalent to maintain their ranking among the high achievers.

Percentiles are important in interpreting gains because they provide the basis for answering the question "Did the student's ranking change?" Grade equivalents, on the other hand, do not answer this question. They answer the question "How much did the student learn?" The grade equivalent scale answers this question in units roughly equivalent to one year's growth for an average (50th percentile) student.

With this distinction between the two scales being clear, the apparent antithesis in growth is more easily understood. High-, average-, and low-achieving students may maintain their various rankings in the population while making different grade equivalent gains. Only at the 50th percentile level would a gain of 1.0 grade equivalent be necessary and sufficient to maintain the same percentile rank.

Figure 6 presents an example of this issue using the Iowa Tests of Basic Skills, 1978, Language Total norms for the spring, grades 3-6. A 25th percentile third-grade student who maintains that ranking across three years of instruction will gain 2.42 grade equivalents compared to a gain of 3.35 for a 75th percentile student. These two students will have maintained their relative rankings; however, the gap between them will have increased by over nine months in three years. To have prevented this gap from widening, the 25th percentile student would have needed an increase from the 25th to the 41st percentile across these three years. For these two students, equal gains in grade equivalents would have resulted in a 16 percentile point greater gain for the lower achiever.

| PERCENTILE | GRADE EQUIVALENT | | | | THREE-YEAR GAIN |
|---|---|---|---|---|---|
| | GRADE 3 | GRADE 4 | GRADE 5 | GRADE 6 | |
| **25** | 2.93 | 3.75 | 4.55 | 5.35 | **2.42** |
| **50** | 3.85 | 4.85 | 5.80 | 6.85 | **3.00** |
| **75** | 4.95 | 6.05 | 7.15 | 8.30 | **3.35** |

Figure 6. Grade equivalent gains made by low, average, and high achievers who maintain the same percentile rank across three years--ITBS, Language Total, spring norms.

b. For individual students, the importance of inspecting both percentiles and grade equivalents when interpreting achievement growth is obvious from the preceeding examples. The same importance is present when considering measures of central tendency for groups. A frequent analysis for public school systems is a comparison of the gap between minority and majority student groups' achievement levels from one school year to another. When one group's median scores are above the 50th percentile and the other's are below, there is real potential for simplistic conclusions which may be misleading.

After a couple of years of reporting that the percentile gap between our minority students and our majority students had been narrowing, we decided to project when the gap would be closed if current trends continued. What we found was that the gap would not ever close. In short, we found that the minorities were gaining less from year to year in terms of grade equivalents than were the majorities. The higher gains in percentiles were an artifact of their relative location in the distribution of scores.

ANOMALY 5: THE SAME GRADE EQUIVALENT EARNS A DIFFERENT PERCENTILE IN EACH CONTENT AREA.
(Six of One is Larger than Half a Dozen of the Other)

So many educators who are surprised when they find that a grade equivalent of 8.2 is the 90th percentile in language but the 96th percentile in math are the same people who say "of course" when someone states that children vary more in their language skills than in their math skills. People can get quite frustrated, however, to find that they cannot straightforwardly compare grade equivalents across content areas.

Figure 7 presents the percentiles associated with certain grade equivalents for the ITBS Language Total and Math Total, grade 5, spring norms. Notice that only at the 50th percentile are the two percentile scales matched up. By definition, they have to match at the 50th percentile. However, since math skills do vary less across students than do language skills, the math percentiles change more slowly as one goes either higher or lower from the 50th percentile.

A student who is at the 90th percentile on both tests receives a grade equivalent of 8.2 on Language Total and a 7.6 on Math Total. This student is farther ahead of the same proportion of peers in both areas, but is farther above grade level in math.

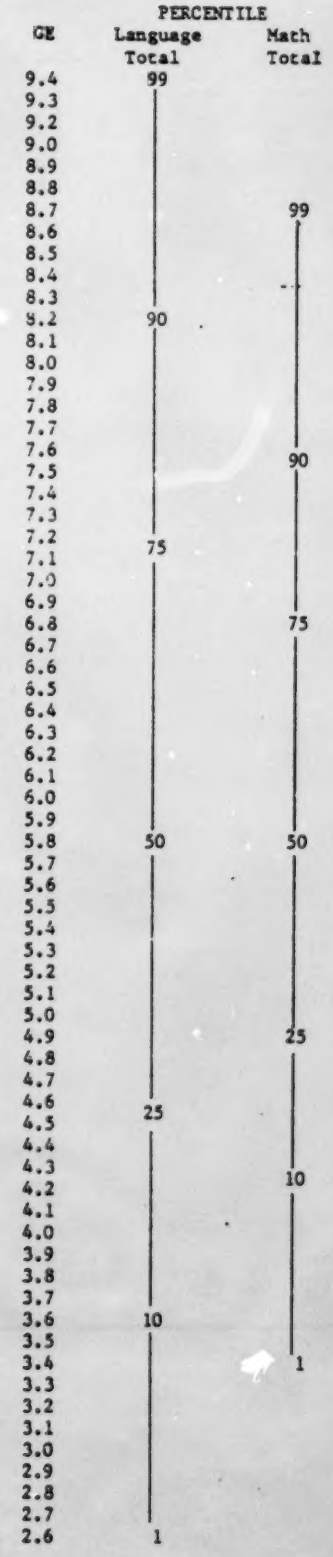


Figure 7. Comparison of grade equivalent and percentile scores--ITBS, Language Total and Math Total, grade 5, spring norms.

## ANOMALY 6: THE MEDIAN DOES NOT REPRESENT ANY GROUP (Nobody Wants to be Considered Average.)

Achievement test scores have a mystique. School personnel may feel uncomfortable talking about them because they do not understand the terms used, such as percentile, grade equivalents, standard error, or normal curve equivalent. Even in districts where intensive efforts have been made to educate the personnel to a few terms, anxiety and misunderstanding still abound. Our office has attempted to ensure that all AISD school personnel understand the basic statistic used in reporting our achievement results-- the median percentile score. We have not been truly successful.

A partial explanation for this inability to understand this "simple" concept may be that it is not, as discovered during examination of our 1981 test results, a simple concept in practice. There are times when the median percentile score for a school does not really represent any single group of students (grouping along traditional lines, like ethnicity). In this situation the score may seem incorrect and meaningless, and school personnel may indeed lose confidence in the utility of the score.

Figure 8 provides a case in which the median percentile score for all students tested does not really represent any one group of students by ethnicity. In reading, the Anglo/Other median was 31 points higher than the total group median, with the medians for Black and Hispanic students 18 and 20 points lower than the total group median percentile score. The third-quartile scores for the Black and Hispanic students are lower than the total group median score, and the first-quartile score for the Anglo/Other group is equal to the total group median percentile score. Thus we have two contrasting groups in terms of achievement, the Anglo/Other and the Black/Hispanic. The total group median percentile is a score which really does not represent any group in the school.

Ethnicity aside, this is a school which has many high achievers and many low achievers and fewer average ones--definitely bimodal. A single school median masks this. Whenever possible, subgroup medians need to be examined prior to using a total group median to describe a school.

| | Total | Black | Hispanic | Anglo/Other |
|---|---|---|---|---|
| Number Tested | 428 | 66 | 132 | 230 |
| Third Quartile | 63 %ile | 21 %ile | 24 %ile | 77 %ile |
| Median | 30 %ile | 12 %ile | 10 %ile | 61 %ile |
| First Quartile | 10 %ile | 5 %ile | 4 %ile | 30 %ile |

Figure 8. 1981 STEP II Percentile Scores, Grade 9, Reading. (Actual School Data)

## Conclusion

In addition to curriculum/test content matches, reliability estimates, and a dozen other issues which confound the straightforward interpretation of achievement test scores, the interaction of types of scores, such as percentiles and grade equivalents, shifts in student demographics, and non-normal distributions within groups being tested must be carefully considered in interpreting achievement test results. This paper describes six anomalies which have been recently encountered by our school system. As long as we evaluators and researchers are called upon to interpret test results, we must be able to distinguish real achievement gains from artifactual gains resulting from anomalies such as these discussed here.